I claim:

1. A method of representing a set of images for pattern classification, the method comprising:

receiving data points corresponding to the set of images in an input space;

generating a neighboring graph indicating whether the data points are neighbors;

estimating geodesic distances between the data points based upon the neighboring graph;

representing each of the data points by an associated feature vector corresponding to the geodesic distances to other data points;

applying Fisher Linear Discriminant to the feature vectors associated with the data points to obtain an optimal direction for projecting the feature vectors for pattern classification.

2. The method of claim 1, wherein generating a neighboring graph comprises:

determining distances between the data points;

determining whether the data points are neighbors based on the determined distances;

responsive to determining that the data points are neighbors, selecting the determined distance for the neighboring graph;

responsive to determining that the data points are not neighbors, selecting an infinite value for the neighboring graph.

3. The method of claim 2, wherein determining whether the data points are neighbors comprises selecting a predetermined number of closest data points from each data point based on the determined distance as the neighbors.

4. The method of claim 2, wherein determining whether the data points are neighbors comprises selecting data points within a predetermined radius from each data point based on the determined distance as the neighbors.

5. The method of claim 1, wherein estimating the geodesic distance between the data points comprises approximating the geodesic distance between the data points with a distance covered by a sequence of short hops between neighboring data points on the neighboring graph using the Floyd-Warshall algorithm.

6. The method of claim 1, wherein applying Fisher Linear Discriminant to the feature vectors comprises projecting the feature vectors to a lower dimensional space lower in dimension than the input space so as to substantially maximize a variance between clusters of feature vectors while substantially minimizing the variance within each cluster of the feature vectors.

7. The method of claim 6, wherein the variance between the clusters of the feature vectors is represented by a between-class scatter matrix and the variance within each cluster of the feature vectors is represented by a within-class scatter matrix.

8. The method of claim 7, wherein the feature vectors are projected to the lower dimensional space so as to substantially maximize a ratio of the between-class scatter matrix to the within-class scatter matrix.

9. The method of claim 1, wherein the images are face images or digit images.

10. A method of representing a set of images for pattern classification, the method comprising:

receiving data points corresponding to the set of images in an input space;

generating a neighboring graph indicating whether the data points are neighbors;

estimating geodesic distances between the data points based upon the neighboring graph;

representing each of the data points by an associated feature vector corresponding to the geodesic distances to other data points;

applying Kernel Fisher Linear Discriminant to the feature vectors associated with the data points to obtain an optimal direction for projecting the feature vectors for pattern classification.

11. The method of claim 10, wherein generating a neighboring graph comprises:

determining distances between the data points;

determining whether the data points are neighbors based on the determined distances;

responsive to determining that the data points are neighbors, selecting the determined distance for the neighboring graph;

responsive to determining that the data points are not neighbors, selecting an infinite value for the neighboring graph.

12. The method of claim 11, wherein determining whether the data points are neighbors comprises selecting a predetermined number of closest data points from each data point based on the determined distance as the neighbors.

13. The method of claim 11, wherein determining whether the data points are neighbors comprises selecting data points within a predetermined radius from each data point based on the determined distance as the neighbors.

14. The method of claim 10, wherein estimating the geodesic distance between the data points comprises approximating the geodesic distance between the data points with a distance covered by a sequence of short hops between neighboring data points on the neighboring graph using the Floyd-Warshall algorithm.

15. The method of claim 10, wherein applying Kernel Fisher Linear Discriminant to the feature vectors comprises:

projecting the feature vectors to a high dimensional feature space using a projection function;

generating Kernel Fisherfaces for the feature vectors projected to the high dimensional feature space;

projecting the feature vectors to a lower dimensional space lower in dimension than the input space and the high dimensional feature space based on the Kernel Fisherfaces so as to substantially maximize a variance between clusters of feature vectors while substantially minimizing the variance within each cluster of the feature vectors.

16. The method of claim 15, wherein the variance between the clusters of the feature vectors is represented by a between-class scatter matrix and the variance within each cluster of the feature vectors is represented by a within-class scatter matrix.

17. The method of claim 16, wherein the feature vectors are projected to the lower dimensional space so as to substantially maximize a ratio of the between-class scatter matrix to the within-class scatter matrix.

18. The method of claim 17, wherein a fraction of an identity matrix is added to the within-class scatter matrix.

19. The method of claim 15, wherein the projection function $\Phi(x)$ satisfies the following relation:

$$k(x, y) = \Phi(x) \bullet \Phi(y)$$

where $k(x,y)$ is a kernel function, $\Phi(x) \bullet \Phi(y)$ is the dot product of the projection functions $\Phi(x)$ and $\Phi(y)$, and x and y are real number variables.

20. The method of claim 10, wherein the images are face images or digit images.

21. A system for representing a set of images for pattern classification, the system comprising:

neighboring graph generation module for receiving data points corresponding to the set of images in an input space and for generating a neighboring graph indicating whether the data points are neighbors;

a geodesic distance estimation module for estimating geodesic distances between the data points based upon the neighboring graph;

a Fisher Linear Discriminant module for representing each of the data points by an associated feature vector corresponding to the geodesic distances to other data points and for applying Fisher Linear Discriminant to the feature vectors associated with the data points to obtain an optimal direction for projecting the feature vectors for pattern classification.

22. The system of claim 21, wherein the neighboring graph generation module generates the neighboring graph by:

determining distances between the data points;

determining whether the data points are neighbors based on the determined distances;

responsive to determining that the data points are neighbors, selecting the determined distance for the neighboring graph;

responsive to determining that the data points are not neighbors, selecting an infinite value for the neighboring graph.

23. The system of claim 22, wherein determining whether the data points are neighbors comprises selecting a predetermined number of closest data points from each data point based on the determined distance as the neighbors.

24. The system of claim 22, wherein determining whether the data points are neighbors comprises selecting data points within a predetermined radius from each data point based on the determined distance as the neighbors.

25. The system of claim 21, wherein the geodesic distance estimation module estimates the geodesic distance between the data points by approximating the geodesic distance between the data points with a distance covered by a sequence of short hops between neighboring data points on the neighboring graph using the Floyd-Warshall algorithm.

26. The system of claim 21, wherein the Fisher Linear Discriminant module applies Fisher Linear Discriminant to the feature vectors by projecting the feature vectors to a lower dimensional space lower in dimension than the input space so as to substantially maximize a variance between clusters of feature vectors while substantially minimizing the variance within each cluster of the feature vectors.

27. The system of claim 26, wherein the variance between the clusters of the feature vectors is represented by a between-class scatter matrix and the variance within each cluster of the feature vectors is represented by a within-class scatter matrix.

28. The system of claim 27, wherein the feature vectors are projected to the lower dimensional space so as to substantially maximize a ratio of the between-class scatter matrix to the within-class scatter matrix.

29. The system of claim 21, wherein the images are face images or digit images.

30. A system for representing a set of images for pattern classification, the system comprising:

neighboring graph generation module for receiving data points corresponding to the set of images in an input space and for generating a neighboring graph indicating whether the data points are neighbors;

a geodesic distance estimation module for estimating geodesic distances between the data points based upon the neighboring graph;

a Kernel Fisher Linear Discriminant module for representing each of the data points by an associated feature vector corresponding to the geodesic distances to other data points and for applying Kernel Fisher Linear Discriminant to the feature vectors associated with

the data points to obtain an optimal direction for projecting the feature vectors for pattern classification.

31. The system of claim 30, wherein the neighboring graph generation module generates the neighboring graph by:

determining distances between the data points;

determining whether the data points are neighbors based on the determined distances;

responsive to determining that the data points are neighbors, selecting the determined distance for the neighboring graph;

responsive to determining that the data points are not neighbors, selecting an infinite value for the neighboring graph.

32. The system of claim 31, wherein determining whether the data points are neighbors comprises selecting a predetermined number of closest data points from each data point based on the determined distance as the neighbors.

33. The system of claim 31, wherein determining whether the data points are neighbors comprises selecting data points within a predetermined radius from each data point based on the determined distance as the neighbors.

34. The system of claim 30, wherein the geodesic distance estimation module estimates the geodesic distance between the data points by approximating the geodesic

distance between the data points with a distance covered by a sequence of short hops between neighboring data points on the neighboring graph using the Floyd-Warshall algorithm.

35. The system of claim 30, wherein the Kernel Fisher Linear Discriminant module applies Kernel Fisher Linear Discriminant to the feature vectors by:

projecting the feature vectors to a high dimensional feature space using a projection function;

generating Kernel Fisherfaces or the feature vectors projected to the high dimensional feature space;

projecting the feature vectors to a lower dimensional space lower in dimension than the input space and the high dimensional feature space based on the Kernel Fisherfaces so as to substantially maximize a variance between clusters of feature vectors while substantially minimizing the variance within each cluster of the feature vectors.

36. The system of claim 35, wherein the variance between the clusters of the feature vectors is represented by a between-class scatter matrix and the variance within each cluster of the feature vectors is represented by a within-class scatter matrix.

37. The system of claim 36, wherein the feature vectors are projected to the lower dimensional space so as to substantially maximize a ratio of the between-class scatter matrix to the within-class scatter matrix.

38. The system of claim 30, wherein the images are face images or digit images.

39. The system of claim 35, wherein the projection function $\Phi(x)$ satisfies the following relation:

$$k(x, y) = \Phi(x) \bullet \Phi(y)$$

where $k(x,y)$ is a kernel function, $\Phi(x) \bullet \Phi(y)$ is the dot product of the projection functions $\Phi(x)$ and $\Phi(y)$, and x and y are real number variables.

40. A computer program product for representing a set of images for pattern classification, the computer program product stored on a computer readable medium and adapted to perform a method comprising:

receiving data points corresponding to the set of images in an input space;

generating a neighboring graph indicating whether the data points are neighbors;

estimating geodesic distances between the data points based upon the neighboring graph;

representing each of the data points by an associated feature vector corresponding to the geodesic distances to other data points;

applying Fisher Linear Discriminant to the feature vectors associated with the data points to obtain an optimal direction for projecting the feature vectors for pattern classification.

41. A computer program product for representing a set of images for pattern classification, the computer program product stored on a computer readable medium and adapted to perform a method comprising:

receiving data points corresponding to the set of images in an input space;

generating a neighboring graph indicating whether the data points are neighbors;

estimating geodesic distances between the data points based upon the neighboring graph;

representing each of the data points by an associated feature vector corresponding to the geodesic distances to other data points;

applying Kernel Fisher Linear Discriminant to the feature vectors associated with the data points to obtain an optimal direction for projecting the feature vectors for pattern classification.

42. A system for representing a set of images for pattern classification, the system comprising:

means for receiving data points corresponding to the set of images in an input space and for generating a neighboring graph indicating whether the data points are neighbors;

means for estimating geodesic distances between the data points based upon the neighboring graph;

means for representing each of the data points by an associated feature vector corresponding to the geodesic distances to other data points

and for applying Fisher Linear Discriminant to the feature vectors associated with the data points to obtain an optimal direction for projecting the feature vectors for pattern classification.

43. A system for representing a set of images for pattern classification, the system comprising:

means for receiving data points corresponding to the set of images in an input space and for generating a neighboring graph indicating whether the data points are neighbors;

means for estimating geodesic distances between the data points based upon the neighboring graph;

means for representing each of the data points by an associated feature vector corresponding to the geodesic distances to other data points and for applying Kernel Fisher Linear Discriminant to the feature vectors associated with the data points to obtain an optimal direction for projecting the feature vectors for pattern classification.